DOCUMENT RESUME

ED 108 628 IR 002 145

AUTHOR	Flournoy, Nancy
TITLE	IMS Version 3 1971-72 Tryout: Initial Report on COMSYS 1 and 2 Computer Operations.
INSTITUTION	Southwest Regional Laboratory for Educational Research and Development, Los Alamitos, Calif.
REPORT NO	SWRL-TN-5-72-31
PUB DATE	27 Apr 72
NOTE	16p.
EDRS PRICE	MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS	Computer Programs; Computers; *Data Bases; Data Processing; *Electronic Data Processing; Information Storage; *Information Systems; Management Systems; *Pilot Projects; Recordkeeping; *Student Records; Tests; Test Scoring Machines
IDENTIFIERS	IMS; *Instructional Management System

ABSTRACT

Instructional Management System (IMS) information cycles from participating schools to the Southwest Regional Laboratory (SWRL) for processing, returning again to the school in the form of summarized reports of pupil data. One segment of this cycle is managed by the IMS pupil data base processing program. The program generates reports from processed IMS pupil test scores and updates the Pupil Data Base Tape at a central computing facility. This paper describes the characteristics of IMS data processed during the first seven weeks of the IMS tryout as interpreted and logged by the managing program. Statistics on the various computer runs, forms submitted, and school response rates are provided. (Author/DGC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED108628

IR

SWRL

SOUTHWEST REGIONAL LABORATORY TECHNICAL NOTE

DATE: April 27, 1972

NO: TN 5-72-31

TITLE: IMS VERSION 3 1971-72 TRYOUT: INITIAL REPORT ON COMSYS 1 AND 2 COMPUTER OPERATIONS

AUTHOR: Nancy Flournoy

ABSTRACT

IMS information cycles from the school to SWRL for processing, returning again to the school in the form of summarized reports of pupil data. One segment of this cycle is managed by the IMS pupil data base processing program. The program generates reports from processed IMS pupil test scores and updates the Pupil Data Base Tape at a central computing facility. This paper describes the characteristics of IMS data processed during the first seven weeks of computer operation as interpreted and logged by the managing program.

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

IR 002 145

This document is intended for internal staff distribution and use. Permission to reprint or quote from this working document, wholly or in part, should be obtained from SWRL, 11300 La Cienega Boulevard, Inglewood, California, 90304.

IMS VERSION 3 1971-72 TRYOUT: INITIAL REPORT ON
COMSYS 1 AND 2 COMPUTER OPERATIONS

In the period of January 1 to February 20, 1972, fourteen IMS runs have been processed successfully through the IMS pupil data base processing program. Of these, three (21.4%) were wholly concerned with maintaining and updating the Pupil Data Base Tape at UCLA's CCN. The remaining eleven "Data Runs" were concerned with storing and reporting pupil scores (Table 1).

Definitions and Data

The term "Class Identification" is used to denote the six classification variables: District, School, Teacher, Program, Grade, and Class. A "CLASS ID" sheet with such data must precede a set of pupil data when scanned.

A "Data Run" is defined by its composition: A group of pupil scores from one or more classes processed as a unit. The first 11 Data Runs contained pupil scores from 23 distinct classes with an average of 8.5 and a range of 3 to 16 classes per Run. The size of a given Data Run is largely determined by the rate and volume of data being received, an effort to limit the scanner operator's processing time to under 2 hours, and a restriction in program processing limiting the number of Data Records accepted to less than 500.

A "Delete" request is made to eliminate a pupil from the Pupil Data Base reports, while an "Update" request adds a new student to the Pupil Data Base. Seventeen Deletes have been processed through Data Run 14, but since many were submitted erroneously due to a misunderstanding of the purpose of the Delete Form (see TN-5-72-03), the average of 1.5 per

Data Run may be expected to decrease. Fourteen Updates were also processed without error.

A "Data Record" is a count of logical records excluding those with Class Identification information which provides one measure of data volume processed. A Delete or an Update request is made with one Data Record. A FYCSP Test is composed of two Data Records, one for each page, and a LMS Test page is composed of from one to ten Data Records; each record contains one pupil's test scores.

Data Run Content

Over 1,000 Criterion Exercises have been processed through Run 14 averaging 98 per Data Run. An average of 112 Data Records have been handled per Data Run with a maximum of 217 processed in Run 1. Thus, the limit on the number of Data Records that IMS programs can handle in a Run has not yet been any real restriction on the volume of data composing a Run.

Tables 2 and 3 show the Class content of each Data Run for LMS and FYCSP programs, respectively, indicating the classes which had not yet submitted data. Individual teachers have submitted data into a maximum of six separate Data Runs. Forty percent of the teachers currently participating in LMS programs and twenty percent of those in FYCSP programs have submitted data.

A State of Control

Unexpected circumstances, as well as clerical errors, delay the IMS data cycle process as they must be traced, explained, and corrected. A

variety of problems affecting the IMS process at the program processing stage are listed under "Comments" in Table 1 and in Table 4. Some of the problems may be expected to occur periodically and their frequency counts will embody variations inherent in the IMS program that are impossible, or economically infeasible to remove. In a controlled system, measurements will vary owing to chance causes only and, therefore, approximately satisfy the concept of randomness. When a State of Control is achieved the mass behavior of measurement variation and the limits within which repeated measures may be expected to fall is predictable through statistical methods. This variation is contrasted with variation due to some assignable and remediable determinant.

As expected in a newly implemented processing system, a State of Control has not been reached and many of the problems detected to date stem from assignable causes. Many deficiencies in IMS programs have been corrected and improvements introduced. Likewise, many of the teachers' clerical mistakes were assignable being due to misunderstanding the required IMS procedure and forms. The apparent decreasing frequency of errors with Runs discussed later could be, in part, due to eliminating such misunderstandings. However, since few teachers have submitted data into over two Data Runs, there is as yet too little data to reliably determine the effect of teachers' experience on the frequency of problems detected, i.e., the proportion of measurement variance assignable to teacher caused errors.

Processing Data Runs

A single Data Run may require more than one computer processing run if complications occur. Such complications are noted under "Comments"

in Table 1 and are followed by "(R)" for rerun. Of the 11 Data Runs, 6 (54.5%) required more than one computer transmission from SWRL's data concentrator to the central computing facility.

On the first attempt to transmit Run 2 data, a 690 Read Error occurred. Transmission was successful the next day. In Run 4, an illegal Class ID (District, School, Teacher Program, Grade, and Class) message and an illegal Program-or-Unit-Number message were printed. More illegal Class ID's appeared in Runs 5, 6, 7, and 12 and in each case the unprocessed class data were pulled from the Data Run. The cause was traced to a misprinting of the Class Identification Sheets such that the teacher's coded ID failed to match the Pupil Data Base codes. Only in Runs 7 and 12 was it necessary to rescan the data without the problem class. In these Runs the error pertained to the last class being processed and hence updating of the Pupil Data Base Tape did not terminate properly. All the original reports were good and mailed from the first Run. The illegal data record in Run 4 was traced to a student's stray mark which was corrected. The entire Run was rescanned and reprocessed by the operator although this was repetitive in terms of reports generated. A Pupil Data Base Maintenance Program has been designed (see TN-5-72-11) to manipulate specified data records individually, but it was not operable until February 25.

In Run 11, the software distorted one class unit's reports. The RJS operator attempted to reprocess the data in the absence of the IMS systems programmer. However, a third Run was necessary with the complicating class unit data pulled until the program could be modified to

handle that unit. Another programming bug was detected in Run 13 when an error supposedly specific to FYCSP data occurred for LMS data. Corrected, the Run was reprocessed.

The cause of each computer rerun was "assignable" and thus, with the elimination of each cause, the IMS process approaches its State of Control.

Interpreting Error Frequency Counts

The errors detected by the IMS programs are only a subset of those known to interfere with IMS data processing. Pupil Data being processed under ComSys 1 or 2 input modes initially preprocessed in the Forms Control Station (FCS) where many types of forms and procedure errors are checked for, corrected, and logged by date and teacher. They are presently being cross-classified by Run Number, a task that should lessen as the flow of IMS data stabilizes.

Some of these errors are distinguishable by type only through clerical preprocessing and, therefore, data on them is limited to within the ComSys 1 and 2 operating systems. Through February 9, 1972, the different types of errors in this category included:

Missing Class ID Sheet; only tests received
Missing Class ID Sheet; Update Form used for ID
Missing Class ID Sheet; Delete Form used as ID
Missing test date on ID Sheet
Original Class Record Sheet received
Group designation unclear on Class Record Sheet
Lines skipped on Class Record Sheet
Criterion Exercises stapled together
Messages written on Criterion Exercise Sheets
Erasures necessary for scanning
Marks requiring darkening for scanning
LMS pupil tests received (stimulus sheets)
One LMS Criterion Exercise Sheet used per pupil
New pupil's name entered on LMS Criterion Exercise Sheet

Under ComSys 3, the above errors must be classified together as records non-acceptable for processing, the only exception being the missing test date on a Class ID Sheet. An error check for this omission is being added to the IMS programs which will estimate the date when missing.

Other forms and procedure errors which are noticed in preprocessing are corrected even though they are checked for in IMS programs. Errors corrected in FCS through February 9 as well as by the IMS program include:

Pupil List numbers unmarked
"Tested" box not marked
"Delete" box not marked
Pupil code omitted on FYCSP Criterion Exercise Sheet

The list of error checks made by IMS programs has been continually revised as use reveals more appropriate checks to be made. These will be reported in subsequent documentation of detection procedures. However some considerations pertinent to earlier documentation remain applicable and are discussed below.

Two categories of error are determined by cause: (1) errors caused outside the processing system by teacher or student (called Teacher Errors) and (2) errors caused within the processing system by software logic and operators' mistakes (called System Errors).

When an error is detected, it is reported, but the data record may either be rejected or passed on for further processing depending on the nature of the error. Thus, there are also two categories of errors distinguished as rejectable and non-rejectable. The bounds for assigning errors to these two categories has been modified as data management procedures adapt to new information.

Since the detection of a Rejectable Error halts further processing, subsequent errors occurring on that record remain undetected.

If errors are to be counted, the processing mus be completed, allowing all errors to be detected before the record is rejected. As is, the errors must be considered in a hierarchy determined by the order in which they are considered by the processing program. Only for the error detected first is the frequency count f_1 now available a true measure of its frequency of occurrence F_1. For the second error detected, on a single Data Record, the count f_2 is a minimum bound of its actual frequence F_2. A maximum of $f_1 + f_2$ occurrences of the second error is possible. Thus, for the nth Rejectable Error with f_n occurrences detected,

$$0 \leq f_n \leq F_n \leq \sum_{i=1}^{I} f_i$$

where f_i are the detected frequencies of error, i=1 to n in order of the errors' consideration by the program, and F_n is the actual number of errors made on the record being processed.

If the probability of multiple errors on a record was sufficiently small, f_n could be assumed equal to F_n. This assumption appears only tenuously reasonable, since it also seems reasonable that the probability of making two errors given one error was made is greater than the average probability of making an error. If this effect is large, and we assume it is not, treating these frequencies f_i as error counts is hazardous. For example, the actual frequencies of Error I (F_i) and Error J (F_j) cannot be compared using their detected frequencies f_i and f_j since,

$$0 \leq f_i \leq F_i \leq \sum_{k=1}^{I} f_k,$$

and

$$0 \leq f_j \leq F_j \leq \sum_{k=1}^{J} f_k,$$

where f_i and f_j are the Ith and Jth Rejectable Errors to be considered. The magnitude of overlap in these intervals will become evident upon examining the upper bounds of the frequencies when further data are collected.

Furthermore, if $f_i=F_i$ a summing over the different types of error does not give a total count of errors made, but rather the total number of Data Records that contained one or more errors. It seems at this point more reasonable to judge the IMS processing system in terms of defective records rather than error frequencies; for this, the hierachical system of detection is appropriate.

Errors Detected In The Pupil Data Base Processing Program

The frequencies of error as detected by IMS programs and summed over the 14 Data Runs (exclusive of Invalid Class Identification previously discussed), are found in Table 4 listed by teacher.

Two types of error were detected with substantially greater frequencies than the others. The most frequent was the teacher's failure to mark a pupil as tested on LMS Criterion Exercise when the item responses clearly indicate that he was. The error is consequential only for a pupil who obtains a perfect score. In this case, no error message is produced, and his score is not recorded, since his entire data record

was blank. Although 57 percent of the 63 errors detected through Run 14 were of this type, over half of them (21) were made by one teacher in his first Data Input. He has submitted subsequent datum, and since notified has not repeated this error. However, even adjusting for him does not change this error's position relative to the other errors.

The second most frequently detected error was caused by teachers submitting Criterion Exercises for a pupil after deleting him from the Pupil Data Base. The most frequent cause of this mistake was found to be teachers who misunderstood the intention of the delete option and were deleting pupils absent the day of the test.

Teacher 01 01 03 submitted scores in several program units for a pupil and finally deleted him. The two errors listed for this teacher under "Delete Requests" revealed a programming deficiency of failing to order pupil data within a given Data Run by date and/or program unit; the student was deleted before his scores were read. Temporarily, only one unit per class was included in a Data Run when Deletes were present. The IMS programs were modified to handle such situations.

Teacher 01 03 02 did not actually make six errors as indicated by the program count. One pupil's stray pencil mark invalidated the program code while another deliberately altered his identification number on one page of an FYCSP Criterion Exercise. The latter caused Missing-Page-Error messages under his real identification and Duplicate-Identification-Error messages under a third pupil's identification.

Over all, the detection of errors has been reasonably infrequent and most of those detected have been assigned to either a programming logic

deficiency or a teacher misunderstanding. As the assignable errors are traced and eliminated, the error detection frequencies do appear to decrease: Thirty-nine of the errors listed in Table 4 were detected in the first four Data Runs while only nine were detected in Runs 11 through 14.

Table 1

IMS DATA PROCESSING RECORD

Runs	Trans-mission Dates	Number of Classes	Number of Data Records	Number of Deletes	Updates	Tests	Comments
Run 1	1/3	16	217	5	4	189	
Run 2	(1/25) 1/26	8	88	3	1	72	Transmission Failure (R)
Run 3	1/27	0	0	0	0	0	Maintenance
Run 4	(1/27) 1/31	(4) 3	(106) 83	0	0	60	Illegal Class ID(R) & Program Unit No.
Run 5	1/27	9	142	0	3	92	2 Illegal Class ID's
Run 6	2/1	10	136	0	2	95	3 Illegal Class ID's
Run 7	2/7	6	77	1	0	50	
Run 8	2/8	0	0	0	0	0	Maintenance
Run 9	2/8	0	0	0	0	0	Maintenance
Run 10	2/8	8	83	0	0	80	
Run 11	(2/8) 2/10	(6) 5	(110) 93.	0	1	(95) 80	Software Error/Trans-mission Failure (R)
Run 12	(2/11) 2/14	(8) 7	53	2	2	50	Illegal Class ID(R)
Run 13	(2/14) 2/16	11	145	1	1	143	Software Error(R)
Run 14	2/14	10	115	5	0	101	
Total	----	93*	1,232	17	14	1,077	
Average**	----	8.5	112	1.5	1.3	97.9	

Twenty-three of the 93 were distinct Classes.

** The average is taken over the eleven Data Runs, rather than the total 14.

Numbers in parenthesis are for attempted but unsuccessful computer runs.

Table 2

DATA PROCESSED FROM LMS PROGRAMS

District	School	Teacher	Run 1	Run 2	Run 4	Run 5	Run 6	Run 7	Run 10	Run 11	Run 12	Run 13	Run 14	No. of Runs with data
01	01	01	X							X	X			3
01	01	02										X	X	2
01	01	03	X											1
01	01	04	X	X			X				X	X		5
01	01	05					X					X	X	3
01	01	06	X	X					X					3
01	02	02			X	X						X		3
01	02	04	XX	XX		XX					X	X		5
01	02	05	X	X										2
01	02	07			X	X	X	X			X			5
01	02	08										X	X	2
01	02	09									X	X		2
01	02	10												0
01	02	11	X	X					X	(X)				4
02	01	01												0
02	01	03												0
02	01	04												0
02	01	05												0
02	01	06					X		X				X	3
02	01	07				X			X	X		X	X	5
02	01	08					X	X	X		X		X	5
02	01	09				X	X	X	X	X	X			6
02	01	10												0
02	02	01										XX		1
02	02	02												0
02	02	03												0
02	03	01												0
02	03	02												0
02	03	03							X				X	2

* Seventeen more teachers coded 02 03 04 through 04 01 04 have not submitted pupil data.

Double Xs (XX) imply data from two different classes with the same Teacher were processed in the same Data Run.

Table 3

DATA PROCESSED FROM FYCSP PROGRAMS

District	School	Teacher	Run 1	Run 2	Run 4	Run 5	Run 6	Run 7	Run 10	Run 11	Run 12	Run 13	Run 14	No. of Runs with data
1	2	1												0
1	2	3												0
1	2	6						XX						1
1	3	1	X	X			X		X	X			X	6
1	3	2			X					X				2
1	3	3				X								1
3	2	1												0
3	3	1												0
3	4	1												0
3	4	4												0
4	1	1												0
4	1	2												0
4	1	3												0
4	2	1												0
4	3	2												0
4	3	4												0
4	3	5												0
4	3	6												0
4	3	8												0
4	5	1												0

Double Xs (XX) imply data from two different classes with the same Teacher were processed in the same Run.

Table 4

ERRORS BY EACH TEACHER AS DETECTED IN THE
PUPIL DATA BASE PROCESSING PROGRAM

District	School	Teacher	DELETE REQUESTS				CRITERION EXERCISES					TOTAL	PERCENT
			In Duplicate*	For Non-Active Pupil	With Scores	And 'Tested' Marked**	Without 'Tested' Marked	For Deleted Pupils	With Same Pupil ID	With Missing Pages	Invalid Program or Unit #		
01	01	01					21					21	.33
01	01	03	1	1				9				11	.17
01	01	04					3	2				5	.08
01	02	02			1	2						3	.05
01	02	08		1								1	.02
01	02	09					2					2	.03
01	02	11					3	1				4	.06
01	03	01			1					2		3	.05
01	03	02							2	3	1	6	.10
02	01	07					1					1	.02
02	01	08					5					5	.08
02	01	09					1					1	.02
TOTAL			1	2	2	2	36	12	2	5	1	63	100%
PERCENT			.02	.03	.03	.03	.57	.19	.03	.08	.02	100%	

* Error check modified to apply only to FYCSP data after this one error was detected. Crossed out cells indicate that the probability of a positive frequency count in that cell is zero, the error being specific to FYCSP or LMS data.

** Error check added to the PDBPP as of 2/18/72.